NISTIR 6453

Algebraic constraints implying monotonicity for cubics

Charles Fenimore and
John Libert,
NIST

Michael Brill,
Sarnoff Corporation

U.S. DEPARTMENT OF COMMERCE
Technology Administration
National Institute of Standards
and Technology
Electricity Division
100 Bureau Drive, Stop 8114
Gaithersburg, MD 20899-8114

NIST

NISTIR 6453

Algebraic constraints implying monotonicity for cubics

Charles Fenimore and
John Libert,
NIST

Michael Brill,
Sarnoff Corporation

U.S. DEPARTMENT OF COMMERCE
Technology Administration
National Institute of Standards
and Technology
Electricity Division
100 Bureau Drive, Stop 8114
Gaithersburg, MD 20899-8114

January 2000

Abstract

While it is straightforward to formulate constraints which ensure a cubic polynomial is monotonic on an interval, such constraints may not be in a form which is suitable for use with standard optimization techniques and software. The MATLAB* [2] package is typical: the required constraints are a series of simultaneous inequalities. In what follows, two simultaneous algebraic inequalities on the coefficients of a cubic polynomial are shown to be necessary and sufficient to assure monotonicity on an interval. This transformation of interval constraints to simultaneous algebraic constraints requires the application of basic logic and analysis. The constraints are applied to a problem arising in analyzing the performance of several video quality measurement models [4].

1. Introduction

The monotonicity of a differentiable function on an interval is equivalent to a condition on its derivative over the interval, namely the derivative must not change sign at any point. For polynomials, this condition over the interval is equivalent to conditions at the end points and inflection points in the interval, namely the derivatives at all such points must have the same sign. In the case of a cubic polynomial the sole inflection point can be expressed in a closed algebraic form and monotonicity is shown to be equivalent to a pair of algebraic inequalities on the coefficients of the polynomial. In this form, monotonicity is readily incorporated as a constraint in classical optimization techniques and available scientific software.

A number of investigators have considered monotonic regression, particularly for splines. Andersson and Elfving [1] have studied monotone interpolation and approximation by cubic splines. Ramsay [3] has considered spline-based monotonic regression. In the case of cubic polynomials, there does not appear to be a basic package for monotonic regression. This paper describes a mathematical formulation of the constraints which enables the use of available software to solve the cubic problem.

This study in low order fitting arises from a quality measurement problem. In recent years, as digital imaging and video technology has replaced analog systems, much attention has been focused on measures of video and image quality. The main new impetus for such attention is the use of lossy compression techniques to reduce data rates. Lossy means data is not recoverable. Such losses are hopefully imperceptible.

In an international comparison of video quality measurement computational models, the computed results are to be compared with the results of subjective tests [4]. The measurements for such subjective assessment may saturate at the extremes of the scale. The design of the test permits non-linear adjustments of the computed results. These "adjustments" include remapping the computed (or "objective") data with (among other functions) a monotonic cubic polynomial which best fits the objective and subjective

* This article refers to commercial products by name in order to specify the means by which the computed results were obtained. Such mention is not meant to imply the product is the best available for the purpose nor is it an endorsement.

scores. A cubic polynomial is included because it is the lowest degree polynomial which can have an inflection point, and can thus compensate for either saturation or steepening at the extremes of the interval. On the other hand, rank ordering of the subjective scores is locally determined and the design of the test expresses confidence in this ranking by imposing monotonicity on the cubic. Doing so avoids reordering the objective scores.

2. Theory

Without loss of generality, we consider the fitting problem over the unit interval [0,1] for a cubic polynomial, p, to data (t_i,d_i). Other intervals can be scaled to the unit interval.

$$p(t) = x_0 + x_1 \cdot t + x_2 \cdot t^2 + x_3 \cdot t^3$$
$$0 \le t \le 1$$

To achieve monotonicity one must avoid a change in sign of the derivative, p'. A necessary, but insufficient, condition is the derivative of p must have the same sign at 0 and 1. That is, the function a defined below must be nonnegative:

$$\mathrm{A}: p'(0)p'(1) = a = x_1 \cdot (x_1 + 2x_2 + 3x_3) \ge 0$$

The location of the single inflection point, t_I, of $p(t)$

$$t_I = -x_2/3x_3$$

is the second piece of information needed to establish monotonicity. When t_I is not in the open unit interval, (0,1), Condition A implies that p is monotonic on [0,1]. This is a consequence of the following Lemma.

Lemma 2.1. Assume the function f is twice continuously differentiable on the interval $[\alpha,\beta]$. If f' has the same sign at α and β and f'' does not change sign in (α,β), then f is monotonic on $[\alpha,\beta]$.

Proof: Assume that $f'' \ge 0$ on (α,β). For any t in (α,β),

$$f'(\alpha) \le f'(t) = f'(\alpha) + \int_\alpha^t f''(s)ds \le f'(\beta).$$

As a consequence the derivative, f', never changes sign on $[\alpha,\beta]$ and f is monotonic. If instead, $f'' \le 0$, $-f$ would satisfy the above conditions and f would be monotonic. ||

Lemma 2.2 shows that in the case that t_I is in (0,1), monotonicity of a cubic is equivalent to p' having the same sign at the inflection point as at the endpoints of the interval.

Lemma 2.2. A cubic polynomial, $p(t)$, is monotonic on the interval [0,1] if and only if

$$A: p'(0)p'(1) = a = x_1 \cdot (x_1 + 2x_2 + 3x_3) \geq 0$$

and

$$B: t_1 \in (0,1) \text{ implies}$$
$$C: (p'(0) + p'(1)) \cdot p'(t_1) \geq 0$$

Proof: A direct calculation shows that monotonicity implies the two conditions.

Monotonicity of p on [0,1] can be established by considering three cases.

I. For $t_1 \notin (0,1)$, the conditions of Lemma 2.1 are met and p is monotonic.

II. For $t_1 \in (0,1)$ and $p'(t_1) = 0$, p' has a zero of order 2 at t_1, $p'(t) = 3 \cdot x_3 \cdot (t - t_1)^2$. Because its derivative does not change sign on the interval, p is monotonic.

III. For $t_1 \in (0,1)$ and $p'(t_1) \neq 0$, the conditions of Lemma 2.2 require that $p'(0)$, $p'(1)$, and $p'(t_1)$ agree in sign. To see this, first observe that Condition A implies that $p'(0)$ and $p'(1)$ have the same sign and as a consequence $p'(0) + p'(1)$ also agrees in sign. Condition C implies that $p'(t_1)$ also has the same sign.

Lemma 2.1 assures that p is monotonic on each of the subintervals $[0,t_1]$ and $[t_1,1]$, p having a single inflection point. The non-zero derivative, $p'(t_1)$, assures that p is monotonic in the same sense on both subintervals and therefore on the entire unit interval. ||

So, two constraints which are necessary and sufficient to assure monotonicity are: A and (B implies C). Now, recast the logical implication of the second constraint in Lemma 2.2 as a single inequality. Since Condition A is already an inequality, doing so exhibits simultaneous inequalities which are equivalent to the interval constraint for monotonicity.

Theorem 2.3: The cubic polynomial, $p(t) = x_0 + x_1 t + x_2 t^2 + x_3 t^3$, is monotonic on the interval [0,1] if and only if

$$x_1 \cdot (x_1 + 2x_2 + 3x_3) \geq 0$$

and

$$\sqrt{(x_2^2 + 3x_2 \cdot x_3)^2 + (2x_1 + 2x_2 + 3x_3)^2 (3x_1 \cdot x_3^2 - x_2^2 \cdot x_3)^2}$$
$$+ x_2^2 + 3x_2 x_3 + (2x_1 + 2x_2 + 3x_3)(3x_1 \cdot x_3^2 - x_2^2 \cdot x_3) \geq 0.$$

***Proof*:** It is sufficient to replace the implication in Lemma 2.2 with the second inequality.

First, recast B and C as polynomial expressions b and c, assuming $x_3 \neq 0$.

$$B: t_1 \in (0,1) \Leftrightarrow \left|\frac{-x_2}{3x_3} - \frac{1}{2}\right| < \frac{1}{2} \Leftrightarrow b = x_2^2 + 3x_2x_3 < 0$$

and

$$C: (p'(0) + p'(1))p'(t_1) = (2x_1 + 2x_2 + 3x_3) \cdot (x_1 - \frac{2x_2^2}{3x_3} + \frac{x_2^2}{3x_3})$$

$$= (2x_1 + 2x_2 + 3x_3)\frac{(3x_1 \cdot x_3^2 - x_2^2 \cdot x_3)}{3x_3^2} \geq 0$$

$$\Leftrightarrow c = (2x_1 + 2x_2 + 3x_3)(3 \cdot x_1 \cdot x_3^2 - x_2^2 \cdot x_3) \geq 0$$

Second, observe that (B implies C) is equivalent to ($b < 0$ implies $c \geq 0$).

- In the case $x_3 \neq 0$, the equivalence is a consequence of the preceding paragraph.
- In the case $x_3 = 0$, B is false so (B implies C) holds vacuously and $b \geq 0$ so ($b < 0$ implies $c \geq 0$) holds vacuously. The statements are equivalent.

Third, convert the previous implication to an inequality. By logical calculus, the statement ($\mathcal{B}$ implies $\mathcal{C}$) is equivalent to ($\mathcal{C}$ or not $\mathcal{B}$) [5 - page7].]. As a consequence, in the plane of vectors (b, c), ($b < 0$ implies $c \geq 0$) is satisfied by vectors with ($b \geq 0$ or $c \geq 0$). This is the entire plane except for the third quadrant, ($b < 0$ and $c < 0$). The vectors are at an angle of more than $\pi/4$ to the vector (-1, -1). Noting that the dot product of two unit vectors is the cosine of the angle between them

$$\frac{(b,c)}{\|(b,c)\|} \bullet \frac{(-1,-1)}{\sqrt{2}} \leq \frac{1}{\sqrt{2}},$$

equivalently,

$$\sqrt{b^2 + c^2} + b + c \geq 0.$$

Replacing b and c with the corresponding polynomials in the coefficients x_j yields the second inequality. ||

Note that the region satisfying the constraints is not convex, but it is the union of convex regions. These regions are defined by algebraic constraints which can be found by factoring the constraints described in Lemma 2.2 and Theorem 2.3.

3. Application

In an international comparison of several video quality measurement models [4], it was required to fit a monotonic cubic polynomial to the output of each model. The Figures are the results of two least squares fits to a typical data set. Figure 1 displays the results of unconstrained fitting. The best fitting cubic is not monotonic on [0,1]. The two critical points are a local maximum at $t \cong 0.82$ and a local minimum at $t \cong 5.28$.

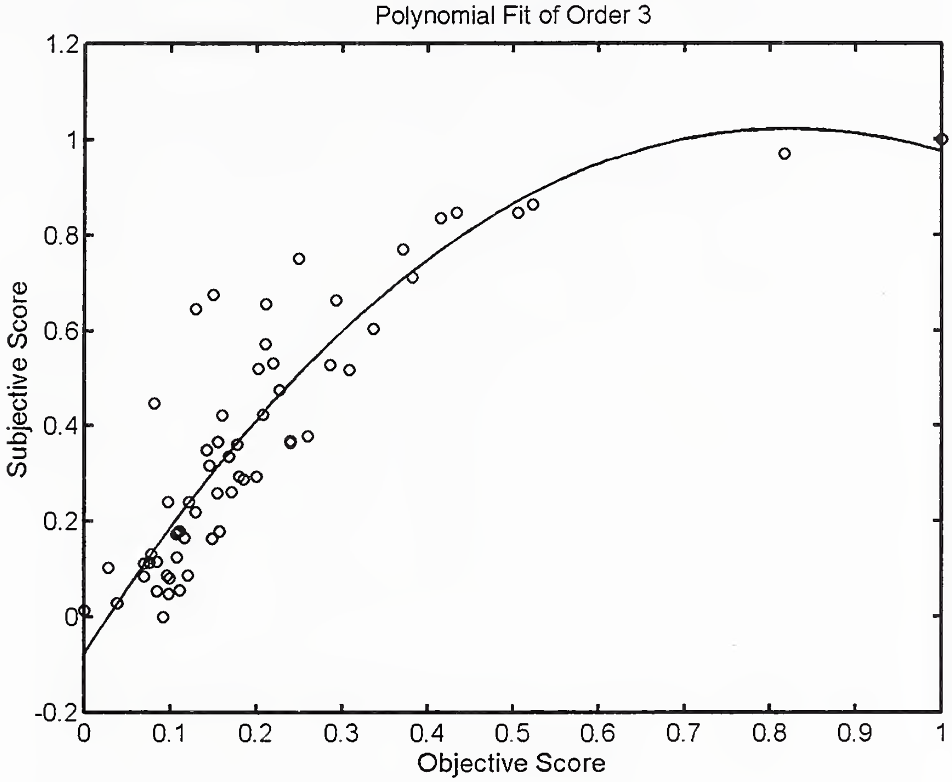

Figure 1: Unconstrained regression of cubic polynomial. The fitted curve is not monotonic Parameter Values: $x_0 = -0.075911$, $x_1 = 2.824086$, $x_2 = -1.99055$, $x_3 = 0.217546$. A local maximum occurs at $t \cong 0.82$ and a local minimum at $t \cong 5.28$.

Figure 2 displays the best fitting cubic which is monotonic on [0,1]. Again, there are two critical points: a local maximum, t ≅1.00, and a local minimum, t ≅1.21. The constrained solution "sweeps" the interior critical point to the boundary of the interval. The results were obtained with the MATLAB Optimization Package [2] using the CONSTR function with the constraints defined here.

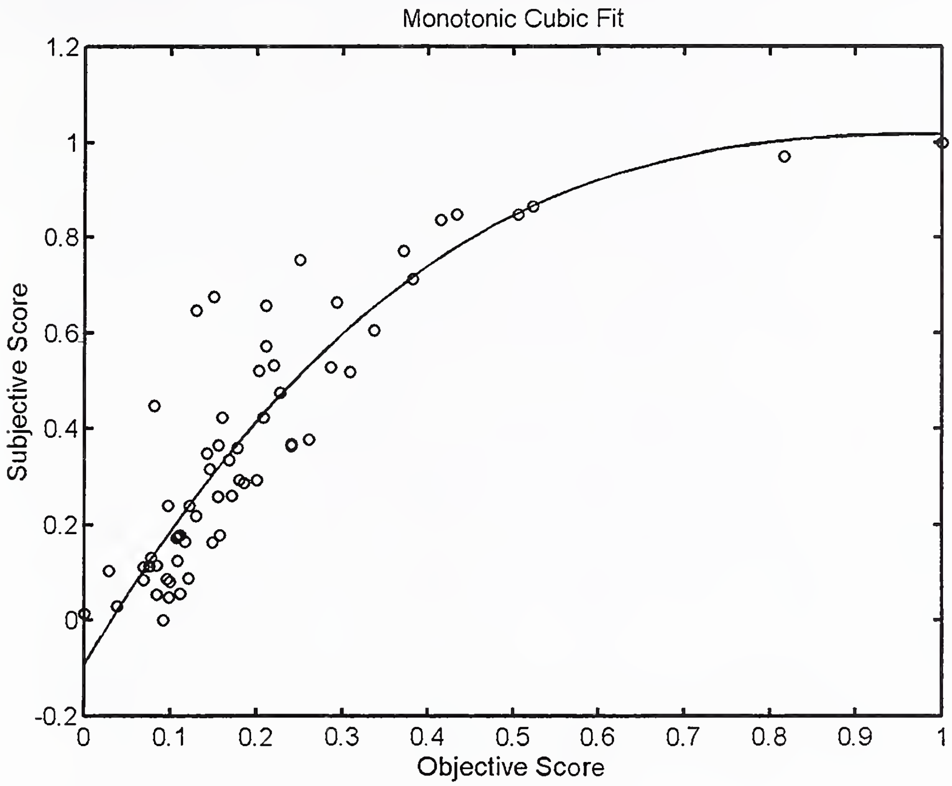

Figure 2: Monotonic regression of cubic polynomial. The fitted curve is monotonic on [0,1]. Parameter Values: x_0 = -0.092407, x_1 = 3.06557, x_2 = -2.7999538, x_3 = 0.844523. Although it is monotonic on [0,1], the cubic has a local maximum at t ≅ 1.00 and a local minimum at t ≅ 1.21.

References

1. L. Andersson and T. Elfving, Interpolation and Approximation by Monotone Cubic Splines, *J. Approx. Theory* **66** (1991) 302-333.
2. MATLAB Constrained Optimization Software, http://www.mathworks.com.
3. J. O. Ramsay, Estimating smooth monotone functions, *J. Royal Statistical Soc. Series B* **60** (1998) 365-375.
4. The Video Quality Experts Group (VQEG), http://www.crc.ca/vqeg/frames.html
5. A.N. Whitehead and B. Russell, "Principia Mathematica", Cambridge U. P., Cambridge, 1967.